Homework 3
posted November 6, 2017
due November 20, 2017

Let $B_n = \{0, 1\}^n$ and let $K \subseteq B_n$. The sequence of Chow parameters of $K$ is $\text{chow}(K) = (c_1, \ldots, c_n, c_K) \in \mathbb{N}^n$ defined as $c_K = |K|$ and $c_i = |\{x \in K \mid x_i = 1\}|$. For example, for $n = 4$ and $K = \{(0, 1, 1, 1), (0, 0, 1, 1), (0, 0, 0, 1)\}$ we have $\text{chow}(K) = (0, 1, 2, 3, 3)$.

Two subsets $K, G$ of $B_n$ are equipollent if they have the same Chow parameters.

The subsets $K$ and $B_n - K$ are linearly separable if there exists a pair $(w, t) \in \mathbb{R}^n \times \mathbb{R}$ such that

$$K = \{x \in B_n \mid w^t x \geq t\} \text{ and } B_n - K = \{x \in B_n \mid w^t x < t\}.$$  

We say that $K$ is linearly separable if $K$ and $B_n - K$ are linearly separable.

1. Let $K \subseteq B_n$. Prove that $\text{chow}(K) = (\sum_{x \in K} x, |K|)$.

2. A diagonal of $B_n$ is a pair $(u, v) \in B_2^n$ such that $u = 1_n - v$, where $1_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$. Prove that if $K$ is a linearly separable subset of $B_n$ that contains a diagonal of $B_n$, then it contains a point of every other diagonal of $B_n$.

3. The optimization problem of the separable data case that seeks to determine a separating hyperplane in $\mathbb{R}^n$ can be transformed into an equivalent optimization problem in $\mathbb{R}^{n+1}$ that seeks to identify a separating subspace. Given a data set $s = ((x_1, y_1), \ldots, (x_m, y_m))$ prove that there exists $r \in \mathbb{R}^n$ such that $s$ is separable by a hyperplane if and only if the set $\bar{s} = ((x_1 + r, y_1), \ldots, (x_m + r, y_m))$ is separable be a subspace $M$ of $\mathbb{R}^n$.

4. Consider the data set $D$ in $\mathbb{R}^2$ shown in Figure 1, where $C$ is a circle centered in $(6, 4)$ having radius 3. Define a transformation $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ such that $\phi(D)$ is linearly separable.

5. There are 16 functions of the form $f : \{0, 1\}^2 \rightarrow \{0, 1\}$. For each such function consider the sequence $S_f = ((x_1, y_1), \ldots, (x_4, y_4))$, where
Figure 1: Non-linearly separable data; positive examples are filled circles.

\[ x_i \in \{0, 1\}^2 \quad \text{and} \quad y_i = \begin{cases} 
-1 & \text{if } f(x_i) = 0, \\
1 & \text{if } f(x_i) = 1
\end{cases} 
\]
for \(1 \leq i \leq 4\).

(a) For how many of these functions is \(S_f\) linearly separable?

(b) Note that for \(f : \{0, 1\}^2 \to \{0, 1\}\) defined by \(f(x) = \min\{x_1, x_2\}\) is linearly separable. For \(\eta = 0.1\) and \(\eta = 0.8\) draw the sequence of weights of the perceptron during the learning process.

6. (optional) Provide a robust implementation in R of the perceptron algorithm that will cycle through a prescribed number of iterations (and, thus, will cope with the case when data is not linearly separable). Map the error count as a function of the number of epochs and run some experiments.